

18. Analyze the security considerations involved in managing Big Data. Propose best practices for data security in a Big Data environment.
 19. Hadoop is a powerful framework for distributed data processing. Explain the benefits and drawbacks of using Hadoop for Big Data Analytics.
 20. Big Data and Artificial Intelligence (AI) are often seen as complementary technologies. Discuss how Big Data Analytics can be used to support AI applications.
-

APRIL/MAY 2024

**DECS43A/GECS43A — BIG DATA
ANALYSIS**

Time : Three hours

Maximum : 75 marks

SECTION A — (10 × 2 = 20 marks)

Answer ALL questions.

1. List two characteristics of Big Data.
2. Define the term "Velocity" as one of the characteristics of Big Data.
3. Briefly explain the difference between Business Intelligence and Big Data Analytics.
4. Briefly mention two types of Big Data Analytics.
5. What are the advantages of using NoSQL databases compared to traditional SQL databases?
6. What does "CAP theorem" stand for in the context of NoSQL databases?
7. Mention two key components of the Hadoop ecosystem.

8. What is the name of the distributed file system used in the Hadoop ecosystem?
9. What does YARN stand for in the context of Big Data Analytics?
10. What are "Mappers" and "Reducers" in the MapReduce programming model?

SECTION B — (5 × 5 = 25 marks)

Answer ALL questions.

11. (a) Describe the challenges associated with working with unstructured data.

Or

- (b) Data projects often involve ethical considerations. Discuss one ethical concern related to Big Data Analytics and propose strategies to mitigate this concern.

12. (a) Explain how Big Data Analytics can be used to gain insights into customer behavior.

Or

- (b) Big Data Analytics often involves a variety of data sources. Describe two techniques used to integrate data from disparate sources.

13. (a) Compare and contrast MongoDB and Apache Cassandra, highlighting their use cases.

Or

- (b) Briefly explain the concept of schema design in traditional relational databases.

14. (a) How does schema design differ in NoSQL databases?

Or

- (b) Explain the key features of HDFS that make it suitable for Big Data storage.

15. (a) Discuss the role of YARN (Yet Another Resource Negotiator) in the Hadoop ecosystem.

Or

- (b) Discuss one limitation of MapReduce and describe an alternative Big Data processing framework that addresses this limitation.

SECTION C — (3 × 10 = 30 marks)

Answer any THREE questions.

16. Discuss the potential impact of Big Data on a specific industry of your choice.

17. Big Data Analytics projects often involve complex data pipelines. Design a basic data pipeline for processing social media data for sentiment analysis.